UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/803,831	03/18/2004	John R. Frank	113744.124 (US2)	5913

23483 7590 09/13/2010

WILMERHALE/BOSTON
60 STATE STREET
BOSTON, MA 02109

EXAMINER	
PHAM, HUNG Q	

ART UNIT	PAPER NUMBER
2159	

NOTIFICATION DATE	DELIVERY MODE
09/13/2010	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

teresa.carvalho@wilmerhale.com
whipusptopairs@wilmerhale.com

Office Action Summary	Application No.	Applicant(s)	
	10/803,831	FRANK, JOHN R.	
	Examiner	Art Unit	
	HUNG Q. PHAM	2159	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *11 August 2010*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-16 and 21-28* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-13 and 21-28* is/are rejected.

7)☒ Claim(s) *14-16* is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date *09/02/10*.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

Continued Examination Under 37 CFR 1.114

A request for continued examination under 37 CFR 1.114, including the fee set forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission filed on 08/11/2010 has been entered.

Response to Arguments

- The rejection under 35 U.S.C. § 112, 2nd paragraph, of claim 26 has been withdrawn in view of the amendment.

- Applicant's arguments with respect to the rejection under 35 U.S.C. § 102(b) or/and 103(a) have been considered but are moot in view of the new ground(s) of rejection.

Claim Rejections - 35 USC § 112

The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

Claims 1 and 3 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite or failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

Regarding amended claim 1, the clause ***the toponyms*** (Line 6) references to other items in the claim. It is unclear what item is being referenced.

Regarding claim 3, the clause, i.e., ***that selected toponym- pair***, references to other items in the claim. It is unclear what item is being referenced.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

This application currently names joint inventors. In considering patentability of the claims under 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of 35 U.S.C. 103(c) and potential 35 U.S.C. 102(e), (f) or (g) prior art under 35 U.S.C. 103(a).

Claims 1, 3-12 and 21-27 are rejected under 35 U.S.C. 103(a) as being unpatentable over Smith et al. [Disambiguating Geographic Names in a Historical Digital Library] in view of Wacholder et al. [Disambiguation of Proper Names in Text] and Bagga et al. [Entity-Based Cross-Document Coreferencing Using the Vector Space Model].

Regarding claims 1 and 27, Smith teaches a method and system comprising:

processing, at a processor, a plurality of toponyms (Smith, Page 6-Lines 13-38) ***wherein each of the toponyms has one or more readings*** (Smith, Page 6-Line 39-→Page 7-Line 7);

for each document within a plurality of documents, identifying geo-textual correlations among readings of the toponyms (As disclosed by Smith, documents in the digital library are scanned for possible proper names and assign the names to PERSON, PLACE or DATE category using simple heuristic methods in Nominator (Smith, Page 6[1], Lines 13-16). The system then attempts to match the names classed as geographic, as well as the uncertain names, against a gazetteer (Smith, Page 6 Lines 33-34). Possible place names are disambiguated based on local context, document context and general world knowledge. In general, if there are explicit disambiguating tags that authors put after place names, e.g., "Lancaster, PA", "Vienna, Austria" and if "Philadelphia" and "Harrisburg" occur in the same paragraph, a reference to "Lancaster" is more likely to be the town in Pennsylvania than to the one in England or Arizona (Smith, Page 6 Line 39-Page 7 Line 1). As disclosed by Wacholder, during the analysis process of Nominator, proper names in documents without personal title or unknown first name such as "Ruth Lake", "Beverly Hills", "Panorama Lake" are assigned low positive scores or zero scores and assigned to PLACE category. Further disambiguation is possible during aggregation across documents by merging if the canonical forms and entity type are identical, e.g., "Ruth Lake" (?PLACE) is merged with "Ruth Lake" (PLACE) (Wacholder, Page 207, Left Column, Line 21-Right Column, Line 12). The teaching of Smith using the Wacholder Nominator as inherited features or obvious features indicate the claimed limitation *for each document within a plurality of documents*, e.g., proper names in the scanned documents in the digital library are analyzed, *identifying geo-textual correlations among readings of the toponyms*, e.g., identifying geographic textual correlations such as

[1] *Page 1 is the first page of the reference.*

the occurrence of “Lancaster” that references to the occurrences of "Philadelphia" and “Harrisburg”); and

selecting one of the plurality of toponyms (The Smith's teaching as discussed above reads on the claimed limitation *selecting one of the plurality of toponyms*, e.g., "Philadelphia" is selected from "Philadelphia", “Harrisburg” and “Lancaster”);

selecting a reading of the one toponym (The Smith's teaching as discussed above reads on the claimed limitation *selecting a reading of the one toponym*, e.g., for the selected "Philadelphia", “Lancaster” as a reading of “Pennsylvania” is selected);

determining a value for a confidence that the selected toponym is associated with the selected reading (The Smith's teaching as discussed above reads on the claimed limitation *determining a value for a confidence that the selected toponym is associated with the selected reading*, e.g., a score is given to “Pennsylvania”, the higher score of “Pennsylvania”, the more likely “Pennsylvania” is a location in Pennsylvania according to the occurrence of “Lancaster”).

The missing of Smith and Wacholder is the claimed limitation ***wherein determining said value involves a mathematical summation over the plurality of documents in which geo-textual correlations were identified that involved that toponym-reading pair***.

Bagga teaches a method of cross-document coreferencing when the same person, place, event is discussed in more than one text source (Bagga, Introduction). Bagga further discloses the formula for calculating the score of a tem, i.e., the weight of a term in a vector of terms (Bagga, Page 81, Right Column-Lines 4-15) to determine the similarity of two document represented by extracted terms vectors. The weight of a term t as taught by Bagga is based on *df*, which is the number of documents in the collection that the term t occurs in. The variable *df* of documents over the collection of document is ***a mathematical summation over the plurality of documents***, and within the collection, geographic textual correlations such as the occurrence of

“Lancaster” that references to the occurrences of "Philadelphia" and “Harrisburg” are identified that ***involved*** the term "Philadelphia" and “Lancaster”.

Therefore, it would have been obvious for one of ordinary skill in the art at the time the invention was made to use the Bagga formula to calculate the score in order to co-reference the same place in more than one text source.

Regarding claim 3, Smith, Wacholder and Bagga, in combination, teach all of the claimed subject matter as discussed above with respect to claim 1, Smith further discloses the step of ***selecting a starting value for the confidence for that selected toponym-pair, and wherein computing value further includes modifying the starting value based on the identified geo-textual correlations within the plurality of documents*** (Smith, Page 7 Lines 7-30).

Regarding claim 4, Smith, Wacholder and Bagga, in combination, teach all of the claimed subject matter as discussed above with respect to claim 3, Smith further discloses the step of ***using a method of uniform priors*** (Smith, Page 7 Lines 7-30).

Regarding claim 5, Smith, Wacholder and Bagga, in combination, teach all of the claimed subject matter as discussed above with respect to claim 1, Smith further discloses the step of ***identifying within the plurality of documents toponyms that have associated geographic locations that are nearby to each other*** (Smith, Page 6 Line 39-Page 7 Line 1).

Regarding claim 6, Smith, Wacholder and Bagga, in combination, teach all of the claimed subject matter as discussed above with respect to claim 1, Smith further discloses the

step of ***identifying spatial correlation among geographic references of toponyms that are in textual proximity*** (Smith, Page 6 Line 39-Page 7 Line 1).

Regarding claim 7, Smith, Wacholder and Bagga, in combination, teach all of the claimed subject matter as discussed above with respect to claim 6, Smith further discloses ***textual proximity means within the same document*** (Smith, Page 6 Line 39-Page 7 Line 1).

Regarding claim 8, Smith, Wacholder and Bagga, in combination, teach all of the claimed subject matter as discussed above with respect to claim 6, Smith further discloses ***textual proximity means within the same document or any document closely linked with said same document*** (Smith, Page 6 Line 39-Page 7 Line 1).

Regarding claim 9, Smith, Wacholder and Bagga, in combination, teach all of the claimed subject matter as discussed above with respect to claim 1, Smith further discloses the step of ***processing the plurality of documents by a named entity tagger prior to identifying the geo-textual correlations*** (Smith, Page 6 Line 39-Page 7 Line 1).

Regarding claim 10, Smith teaches a method comprising:

generating, at a processor, information for ranking a document that includes a plurality of toponyms for which there is a corresponding plurality of toponym-place pairs (Smith, Page 6-Lines 13-38),

wherein the place of each toponym-place pair of the plurality of toponym-place pairs identifies a geographical location or region designated by the toponym (Smith, Page 6-Line 39→Page 7-Line 6);

for a selected toponym-place pair of the plurality of toponym-place pairs that is found within the target document (As disclosed by Smith, documents in the digital library are scanned for possible proper names and assign the names to PERSON, PLACE or DATE category using simple heuristic

methods in Nominator (Smith, Page 6[2], Lines 13-16). The system then attempts to match the names classed as geographic, as well as the uncertain names, against a gazetteer (Smith, Page 6 Lines 33-34). Possible place names are disambiguated based on local context, document context and general world knowledge. In general, if there are explicit disambiguating tags that authors put after place names, e.g., “Lancaster, PA”, “Vienna, Austria” and if "Philadelphia" and “Harrisburg” occur in the same paragraph, a reference to “Lancaster” is more likely to be the town in Pennsylvania than to the one in England or Arizona (Smith, Page 6 Line 39-Page 7 Line 1). As disclosed by Wacholder, during the analysis process of Nominator, proper names in documents without personal title or unknown first name such as “Ruth Lake”, “Beverly Hills”, “Panorama Lake” are assigned low positive scores or zero scores and assigned to PLACE category. Further disambiguation is possible during aggregation across documents by merging if the canonical forms and entity type are identical, e.g., <“Ruth Lake” (?PLACE)> is merged with <“Ruth Lake” (PLACE)> (Wacholder, Page 207, Left Column, Line 21→Right Column, Line 12). In a typical document, a single entity may be referred to by many name variants which differ in their degree of potential ambiguity. For example, “Paris” and “Washington” are highly ambiguous out of context but in well edited text they are often disambiguated by the occurrence of a single unambiguous variant in the same document. Thus, “Washington” is likely to co-occur with either “President Washington” or “Washington, D.C.”, but not with both (Wacholder, Page 206, Right Column, Lines 37-51). The teaching of Smith using the Nominator as inherited features or obvious features read on the claimed limitation *for a selected (toponym,place) pair of the plurality of (toponym,place) pairs that is found within the target document*, e.g., <“Washington”, (?PLACE)> in PLACE category of a target document is selected);

---

[2] *Page 1 is the first page of the reference.*

obtaining a pre-computed number for a value of a confidence that the toponym of the selected toponym-place pair refers to the place of the selected toponym-place pair (As discussed above, <"Washington", (?PLACE)> is assigned a low positive score or zero score);

determining if another toponym is present within the target document that has an associated place that is geographically related to the place referred to by the selected toponym-place pair (As discussed above, "Washington, D.C." is determined that has "D.C." that is geographically related to the place referred to by <"Washington", (?PLACE)>); and

if the other toponym is identified within the target document that has an associated place that is geographically related to the place referred to by the selected toponym-place pair, boosting the value of the confidence for the selected toponym-place pair for the target document (As discussed above, <"Washington, D.C."> is identified within the document that has "D.C." that is geographically related to (?PLACE) referred to be <"Washington", (?PLACE)>, the low positive score or zero score of <"Washington", (?PLACE)> is replaced by the more positive score of <"Washington, D.C.">).

Smith and Wacholder do not explicitly teach ***pre-computed number derived from a statistical observation about a large corpus of documents***.

Bagga teaches a method of cross-document coreferencing when the same person, place, event is discussed in more than one text source (Bagga, Introduction). Bagga further discloses the formula for calculating the score of a tem, i.e., the weight of a term in a vector of terms (Bagga, Page 81, Right Column-Lines 4-15) to determine the similarity of two document represented by extracted terms vectors. The weight of a term t as taught by Bagga is based on *df*, which is the number of documents in the collection that the term t occurs in. The variable *df* of documents over the collection of document is ***pre-computed number derived from a statistical observation about a large corpus of documents***.

Therefore, it would have been obvious for one of ordinary skill in the art at the time the invention was made to use the Bagga formula to calculate the score in order to co-reference the same place in more than one text source.

Regarding claim 11, Smith, Wacholder and Bagga, in combination, teach all of the claimed subject matter as discussed above with respect to claim 10, Smith further discloses the step of ***identifying the other toponym based, at least in part, on the other toponym having an associated geographic region that encompasses the place referred to by the selected toponym- place pair*** (Smith, Page 6 Line 39-Page 7 Line 1).

Regarding claim 12, Smith, Wacholder and Bagga, in combination, teach all of the claimed subject matter as discussed above with respect to claim 10, Smith further discloses the step of ***identifying the other toponym based, at least in part, on the other toponym having an associated place that is geographically nearby the place referred to by the selected toponym- place pair*** (Smith, Page 6 Line 39-Page 7 Line 1).

Regarding claim 21, Smith, Wacholder and Bagga, in combination, teach all of the claimed subject matter as discussed above with respect to claim 1, Smith and Wacholder further discloses the step of ***computing at least one confidence value for each reading of that selected toponym*** (Smith, Page 6 Line 39-Page 7 Line 1).

Regarding claim 22, Smith, Wacholder and Bagga, in combination, teach all of the claimed subject matter as discussed above with respect to claim 1, Smith and Wacholder further discloses the step of ***selecting for each toponym among the plurality of toponyms; selecting a respective***

reading for each of the respective selected toponyms and for each selected toponym-reading pair, computing a respective confidence that the each respective selected toponym means that respective selected reading (Smith, Page 6 Line 39-Page 7 Line 1).

Regarding claim 23, Smith, Wacholder and Bagga, in combination, teach all of the claimed subject matter as discussed above with respect to claim1, Smith further discloses that ***a reading of a toponym is a geographical location or region designated by the toponym*** (Smith, Page 6 Line 13-Page 7 Line 1).

Regarding claim 24, Smith, Wacholder and Bagga, in combination, teach all of the claimed subject matter as discussed above with respect to claim 1, Wacholder further discloses that ***computing said value is done iteratively to arrive at the value for the confidence that the selected toponym means that selected reading*** (Wacholder, Page 207, Left Column, Lines 45-59).

Regarding claim 25, Smith, Wacholder and Bagga, in combination, teach all of the claimed subject matter as discussed above with respect to claim 1, Smith further discloses ***the mathematical summation is of previously determined confidences*** (Smith, Pages 7-8, EVALUATION).

Regarding claim 26, Smith, Wacholder and Bagga, in combination, teach all of the claimed subject matter as discussed above with respect to claim 1, Smith further discloses ***the associated place is different from the place referred to by the selected(toponym- place pair*** (Smith, Page 6-Line 39-Page 7-Line 7).

Claims 2 and 28 are rejected under 35 U.S.C. 103(a) as being unpatentable over Smith et al. [Disambiguating Geographic Names in a Historical Digital Library], Wacholder et al. [Disambiguation of Proper Names in Text], Bagga et al. [Entity-Based Cross-Document Coreferencing Using the Vector Space Model] and further in view of Frank et al. [WO 01/63479 A1].

Regarding claims 2 and 28, Smith, Wacholder and Bagga, in combination, teach all of the claimed subject matter as discussed above with respect to claims 1 and 27, but does not teach the step of ***using the value for the confidences generated for the selected toponym-reading pair to rank documents according to their relevance to a search query***.

Frank teach the step of ***using the value for the confidences generated for the selected toponym-reading pair to rank documents according to their relevance to a search query*** (Frank, Page 32 Line 28-Page 33 Line 19).

Therefore, it would have been obvious for one of ordinary skill in the art at the time the invention was made to include the step of ranking as taught by Frank in order to search for a particular document with spatial criteria.

Claim 13 is rejected under 35 U.S.C. 103(a) as being unpatentable over Smith et al. [Disambiguating Geographic Names in a Historical Digital Library], Wacholder et al. [Disambiguation of Proper Names in Text], Bagga et al. [Entity-Based Cross-Document Coreferencing Using the Vector Space Model] and further in view of Naughton [USP 6,240,425 B1].

Regarding claim 13, Smith, Wacholder and Bagga, in combination, teach all of the claimed subject matter as discussed above with respect to claim 12, but not teach the step of ***computing a geographical distance between the place associated with the identified toponym and the place referred to by the selected (toponym,place) pair***.

Naughton teaches the technique of computing a geographical distance between two areas (Naughton, Col. 5 Lines 8-27).

Therefore, it would have been obvious for one of ordinary skill in the art at the time the invention was made to use the step of computing distance as taught by Naughton in Smith method in order to disambiguating geographic names in a document.

Allowable Subject Matter

Claims 14-16 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

Contact Information

Any inquiry concerning this communication or earlier communications from the examiner should be directed to HUNG Q. PHAM whose telephone number is 571-272-4040. The examiner can normally be reached on Monday-Friday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, JAMES K. TRUJILLO can be reached on 571-272-3677. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications

may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

HUNG Q. PHAM
Primary Examiner
Art Unit 2159

/HUNG Q. PHAM/
Primary Examiner, Art Unit 2159
August 03, 2010